Pricing via Functional Size

A Case Study of a Company's Portfolio of 77 Outsourced Projects

Hennie Huijgens
Delft University of Technology and
Goverdson, The Netherlands
h.k.m.huijgens@tudelft.nl

Georgios Gousios
Radboud University Nijmegen
The Netherlands
g.gousios@cs.ru.nl

Arie van Deursen
Delft University of Technology
The Netherlands
arie.vandeursen@tudelft.nl

***Abstract*—A medium-sized west-European telecom company experienced a worsening trend in performance, indicating that the organization did not learn from history, in combination with much time and energy spent on preparation and review of project proposals. In order to create more transparency in the supplier proposal process a pilot was started on Functional Size Measurement pricing (FSM-pricing). In this paper we evaluate the implementation of FSM-pricing in the software engineering domain of the company, as an instrument useful in the context of software management and supplier proposal pricing. We analyzed 77 finalized software engineering projects, covering 14 million Euro project cost and a project portfolio size of more than 5,000 function points. We found that a statistical, evidence-based pricing approach for software engineering, as a single instrument (without a connection with expert judgment), can be used in the subject companies to create cost transparency and performance management of software project portfolios.**

Keywords—Software Economics; Software Pricing; Functional Size Measurement; FSM-pricing; Function Point Analysis.

I. Introduction

This story is about a company that experiences two problems in its software engineering outsourcing. First, a worsening trend is seen in project cost per Function Point (FP), indicating that the organization does not learn from historic projects. Second, much time and energy is spent on preparation and review of fixed price project proposals. Our case study explores whether a new project pricing method helps to solve these problems.

A. Problem Statement

To arrive at a price that is acceptable for both parties involved, most companies rely heavily on expert judgment [1]; where the advice of knowledgeable staff is solicited [2]. Usually this is performed as a bottom up approach, where component tasks are identified and sized and then these individual estimates are aggregated to produce an overall estimate [2].

Yet, in practice effort and/or schedule overruns are business-as-usual [3], despite involvement of experts. Software development is characterized by high cost and schedule overruns [4]. Estimation errors are reported to be essential causes of poor management, due to lack of a solid baseline of size [5].

An alternative method for software project estimation is based on algorithmic cost models (COCOMO 2 is a well-known example) which take cost drivers representing certain characteristics of the target system and the implementation environment and use them to predict estimated effort [2]. In many of these statistical approaches size is assumed to be a key factor to estimate project cost [6] [7]. Usually size of software engineering projects is measured with a formal Functional Size Measurement (FSM) standard [8]. FSM is a method to measure the size of software engineering projects by means of the functionality delivered to users [7], which lays the foundation for a statistical method of project pricing based on functional size. Advantages of such a statistical method are that this will help to improve transparency of estimations and that it can be a good instrument to create continuous improvement of project performance.

Although we did not find evidence in existing literature, our observation in industry is that a purely statistical method is almost never used. If statistical analysis is used, this is usually supplementary to an expert judgment-based approach [1]. And practice shows that in most cases the expert opinion – in many cases supported by reasoning by analogy – is leading when it comes to decision making [9].

B. Research Objectives

The goal of this paper is to examine whether a purely statistical approach to pricing is effective in an outsourcing context. We define an approach to be effective when a so-called win-win situation is achieved: meaning that both involved parties are satisfied and project proposals are perceived to be transparent for all stakeholders. The supplier delivers a service for a price that is higher than the cost, and the customer gets higher value than the paid price. In addition to that the outsourcing context asks for a long-term (5 year) relation. For this purpose we define three research questions:

RQ1: To what extent are both parties involved in the case study satisfied with FSM-pricing?

RQ2: To what extent does FSM-pricing help to improve transparency of project proposals?

RQ3: To what extent does FSM-pricing help to create cost and time improvements?

C. Context

In order answer these research questions, we performed a case study on the implementation and evaluation of FSM-pricing as a single instrument for software management, in a telecom company in a west-European country (in this paper indicated as Company C), and the pricing approach agreed with its main Indian IT-supplier (indicated as Supplier S). We studied data collected from 77 software projects that finalized during a period from 2012 to 2014. Moreover, we conducted 25 interviews including structured as well as open-ended questions. Our study is primarily descriptive, and not comparative: we don't have data to see how other pricing approaches might have worked. Yet, we provide a rigorous analysis of what worked well, and what did not work well using FSM as an instrument for pricing.

The innovation of our study is that we raise the question to what extent a single, statistical, empirical approach to project estimation can reach the goal of transparent project proposals and due to that, cost and time improvements.

The case study shows that FSM-pricing can successfully be used in the practice of COMPANY C and SUPPLIER S, as a statistical, evidence-based pricing approach for software engineering project proposals (RQ1), that FSM-pricing, in both subject companies leads to an improved transparency of project proposals and satisfied stakeholders (RQ2). Furthermore we found that FSM-pricing in our case study does lead on short term to cost improvements, but that no time improvements are realized within both subject companies: average project duration shortens, but average project size gets smaller too (RQ3). Due to the limited scope of the study it is too early to generalize the above mentioned findings to other companies and suppliers of software projects, yet we believe the outcome can help software companies to setup transparent and improving project pricing strategies.

We base the reporting structure of this case study on the linear-analytic structure as described in [10]. In Section II, we survey earlier research on software pricing and discuss the background of FSM-Pricing. In Section III, we chalk out the case study design. In Section IV, we present results and we evaluate validity. In Section V we discuss the results and Section VI includes conclusions and future work.

II. RELATED WORK

When it comes to software pricing, two types of estimation techniques are distinguished to discover the cost of producing a software system; experience-based techniques such as expert judgment and algorithmic cost modeling where cost is estimated as a mathematical function of product, project and process attributes. A well-known example of the latter is Boehm's COCOMO 2 [11]; more methods based on algorithmic software cost models with specific regression formula are widely used in industry, such as the Putnam Model [12], and SEER-SEM [13].

Studies covered in a review by Moløkken and Jørgensen on Surveys on Software Effort Estimation [3] mention a variety of estimation aids; such as work breakdown structure, Functional Size Measurement such as Function Point Analysis (FPA) [7], parametric tools [14], and qualitative methods [15].

For a long time researchers and practitioners have been investigating the use of statistics in software estimation. Since the 90's a limited number of studies has been published on the subject of pricing of projects based on statistics [16] [17]. Despite all models and practices actual software estimation seems difficult. Moløkken and Jørgensen [3] observe that 60-80% of the projects encounter effort and/or schedule overruns. Estimation methods in most frequent use are expert based: expert consultation, intuition and experience, and analogy. Frequent use of expert judgment is advocated because of a lack of evidence that formal estimation models lead to more accurate estimates [3].

Although research in the field of software engineering often shows conclusion instability (where what is true for project one, does not hold for project two) [18], and expert judgement is common practice, studies do emphasize pitfalls. Jørgensen and Gruske [19] argue that estimation professionals in many cases do not use lessons learned from finalized projects. Valerdi [20] mentions cognitive bias that can make experts produce poor estimates. Passos et al. [21] show that many experts generalize from their first estimates to future ones. Recent literature study on agile metrics shows high popularity of velocity for effort estimates in industrial agile teams [22]; yet, cost metrics and size related metrics, and especially metrics related to pricing of projects, are not mentioned. Fink and Lichtenstein [23] address the gap between project size (although measured here in cost and not in functional size) in the software engineering literature and the attention it receives in software contracting research. Madachy et al. [24] argue that due to impreciseness of general software cost parameters such as size, effort distribution, and productivity cost database better are segmented by domain.

Abran et al. [25] uses a FSM-based model to assess productivity and to estimate new projects on fixed and partly variable costs. Ramasubbu et al. [26] [27] reveal complex tradeoffs in choosing configurational choices that are optimized for productivity, quality, and/or profits. A discussion on model-based versus judgment-based is described in [28], indicating a substantial overlap between the two approaches, but also some mismatches.

We did not find studies that describe dedicated use of algorithmic cost models in practice, without interference of expert-judgment based methods. Very limited research is performed specifically on the topic of pricing software projects. We have not found any studies that emphasize the use of FSM as a single instrument for pricing. This is remarkable; several studies on FSM stress that software size is a primary predictor of project effort and thus project cost [7] [6].

III. CASE STUDY DESIGN

A. Theory

1) FSM and FPA

FSM is an industry standard to measure size of software engineering activities. Five FSM methods are certified by ISO as an international standard; in our study IFPUG FPA (ISO 2003c) [8] is used. FSM origins from FPA, designed by Albrecht in 1979 [29] to estimate size of software delivery by means of user functionality. FSM is based on the complete set of functional requirements of a software project. An extensive overview of FSM can be found in [7].

2) FSM-pricing

FSM-pricing, as used in the context of this case study, is a method that we developed for pricing of proposals for software projects to be performed within COMPANY C, by SUPPLIER S. In order to define a fixed price for a project, first FSM is performed to measure the functional size of a project, second the price of the project is determined based on a power trend that is built on historic data of finalized software projects. In our case study we only used historic data of projects that were finalized within the practice of COMPANY C and SUPPLIER S itself. The FSM-pricing method is explained more in detail in paragraph III.E.

B. Research Questions

In the period prior to FSM-pricing become operational within COMPANY C, we discovered two major disadvantages in the current expert-judgment-based estimation approach through analysis of finalized software engineering projects. First, COMPANY C showed a worsening trend in project cost per FP, indicating that the organization did not learn from historic project data. Second, much time and energy was spent on preparation and review of fixed price project proposals, leading to long project durations. To turn the tide on the worsening cost and time performance, and to smoothen the proposal process, a decision was made to change towards an empirical, evidence-based, and analytical way of preparing fixed price project proposals. FSM-pricing was born, having two goals, defined by COMPANY C's management: 1) improve transparency of proposals, and 2) create ongoing cost and time improvements of software delivery due to the expected improved clarity in the delivery process (e.g. less discussion on cost and scope).

Based on this we defined three research questions, with the intention to find out to what extend stakeholders involved in FSM-pricing are satisfied about the method, to what extend the method helps to improve transparency of project proposals, and to what extent cost and time improvements are realized.

C. Case and Subject Selection

FSM-pricing, as described in this paper, was implemented in the software project department of COMPANY C, as part of a transformation program that includes a change from one large European IT-supplier to a large Indian IT-company (SUPPLIER S) for the majority of its software engineering activities for the Customer Relationship Management (CRM), Billing, and Data Warehouse (DWH) applications. Besides the fact that a 5-year sourcing contract was agreed between COMPANY C and SUPPLIER S, both companies were not in any way - besides contractually - related. FSM-pricing aims to implement FSM based on FPA [8] as an approach to improve the capability of the company to challenge SUPPLIER S's proposals for to-be-started software engineering activities. All proposals were fixed-price; no extra time-material cost were allowed unless the scope of a project (in FPs) was changed during the delivery period.

Based on this organizational definition, and driven by the goal to investigate a representative subset of mutually highly different software projects within a company's software portfolio as a whole, we decided to select all software projects to be finalized during the period January 2014 to December 2014, within the business domains CRM, Billing, and DWH of COMPANY C, with SUPPLIER S acting as the main supplier, to be subject of our case study. For benchmarking purposes we used a subset of historic software projects that were finalized in the period 2012 to 2013, within the three business domains of COMPANY C, yet performed by other external suppliers than SUPPLIER S.

D. Data Collection procedures

Data of all software projects that are collected are measured by a team of COMPANY C, supported by measurement specialists of SUPPLIER S. One of the authors of this study was leading COMPANY C'S measurement team during the case study. As a source for the project data we use the formal project administration. All project data is reviewed by the applicable project manager and the financial controller of COMPANY C, and adjusted where needed. We collect both quantitative data (e.g. core metrics such as size, effort, cost, duration) and qualitative data (e.g. project backgrounds, factors that influenced a project) in a measurement repository. Projects cover a mix of the business domains CRM, Billing, and DWH, project types (e.g. newly built systems, enhancements, off-the-shelf packages), and project sizes (e.g. small enhancements, large once-only projects). In all projects the design, build, and testing activities are performed by one or more external suppliers. Most software projects are combined in releases and delivered at one moment to the business organization; each year eight releases are rolled out under guidance of a portfolio management team of COMPANY C.

We collect data on finalized software engineering projects only; stopped or failed projects are not included in our case study. We exclude projects that are only about infrastructure, or that include only non-functional requirements (e.g. performance, security), because these were not to be counted in FPs.

For all to-be-analyzed software engineering projects, we measure project size in Function Points (FPs), according to FSM ISO/IEC 20926 guidelines [8]. FPA is performed by specialists either from a COMPANY C measurement team (in the period that SUPPLIER S is not in scope as main supplier yet), or by a SUPPLIER S measurement team (once SUPPLIER S is in scope as main supplier they perform all FPAs). Every FPA is reviewed on correct utilization of counting practices by an experienced IT-metrics expert who is also one of the authors of this paper, and on correct interpretation of requirements by an applicable subject matter expert of COMPANY C.

E. Analysis Procedure

In order to test whether cost or time improvements are realized we calculate the following performance indicators for each project (we opted for this set of indicators because they were included in the standard set of KPIs within COMPANY C and therefor to be expected as known by both parties management):

1. *Project cost per FP*; total project cost divided by the project size, expressed in Euros/FP;
2. *Build & Test cost per FP*; cost of the Build & Test phase divided by the project size, in Euros/FP;
3. *Project duration per FP*; duration of the project from start of the Initiation phase to technical go live divided by the project size, in Days/FP.
4. *Build & Test duration per FP*; duration of the Build & Test phase divided by the project size, in Days/FP.

When in this study cost per FP or duration per FP is mentioned without any prefix, the project version of each indicator is meant, instead of the Build & Test version. For analysis purposes results of individual projects are aggregated to company level, where project size (FPs) is used as weighting factor. All data used in the analysis were shared and thoroughly reviewed by measurement experts of both COMPANY C and SUPPLIER S.

Based on analysis of projects performed by SUPPLIER S, we calculated two domain-specific baselines on build & test cost per FP; these were going to be the trend lines for FSM-pricing. To create the baseline, we obtained the best fit after conducting a log-log transform. After performing a power regression, the resulting price calculation formula is:

$$Price = \alpha \times (FP)^{\beta} \quad (1)$$

The coefficients α and β may differ per application domain. In the portfolio under study, we typically have $\beta \approx 0.75$. Note that this formula is in line with COCOMO 2's effort estimation formula (which uses KLOC instead of function points) [11]. We use simple regression on project size and build & test cost with power fit. Our foundation of this argument is that such a model facilitates greater analyzability and thus helps improving transparency. For a statistics-based explanation we create a cross correlation table to determine, and filter the strongly dependent variables in our sample out from the regression model. We found that size and duration are all pair-wise highly correlated; we rejected duration and only used size as a predictor for cost. See the technical report for more details on statistics [30].

We prepared two baselines: 1) CRM/Billing ($R^2 = 0.5621$) and 2) DWH ($R^2 = 0.9048$). CRM/Billing domain projects are combined in one baseline because the analysis shows no large differences between projects from both domains, many projects overlap domain borders, and because not enough data were available for proper individual trend lines for both domains. A separate DWH baseline was setup because these projects show a different pattern. See the Technical Report [30] for plotter charts and details on the setup of both baselines.

Based on both baselines a tool was set up for cost calculation in project proposals by SUPPLIER S. For all to be started software projects the fixed price is calculated with this tool. Once the size of a project is counted and reviewed, the tool calculates the price for a project to be performed by SUPPLIER S based on the applicable domain baseline.

Stakeholders from COMPANY C opted strongly for a single pricing approach (only based on statistics), because ongoing discussions on project estimates were expected due to a variety of expert opinions if two approaches were to be used simultaneously, and because of that longer project durations. To reassure stakeholders of SUPPLIER S with doubts on this single method for supplier proposal pricing, a six month's FSM-pricing pilot was started. This pilot is the subject of the case study that is discussed in this paper. Quantitative analysis is performed over the scope of the six-month pilot and the following six months operational use of FSM-pricing.

F. Model Validation Procedure

In order to validate the FSM-pricing method we use a mixed methods methodology, as we are examining a phenomenon with multiple (qualitative and quantitative) tools. We perform a single-case, holistic case study that involves two instruments; a survey consisting of open and closed questions, and a quantitative analysis of actual project data. The survey is performed six months after the start of the case study, the quantitative analysis is performed at the end of the case study period of one year.

To answer RQ1 (To what extent are both parties involved in the case study satisfied with FSM-pricing?) and RQ2 (To what extent does FSM-pricing help to improve transparency of project proposals?) we create a combined 10-minute questionnaire survey. The survey topics and the survey approach are determined in a number of preparation sessions between management representatives and the measurement experts of both COMPANY C and SUPPLIER S. Our aim is to come up with a manageable set of topics that would represent the pilot effectively. The survey consists of a number of closed questions; respondents are asked to rate these survey topics on a 5-point Likert scale. Next to the 5-point scale for each of the survey topics a choice of "Don't Know" as an answer is an option. Besides that the survey contains three open questions.

The survey starts with the collection of demographic information, and the answering of two partially closed questions: "What company are you working for?" and "What is your connection with FSM-pricing?" Both questions are intended to find out any differences in satisfaction with FSM-pricing within both the involved parties COMPANY C and SUPPLIER S, and between respondents with different roles. A comprehensive overview of setup and respondent statements in the survey can be found in the technical report [30].

To assess the experienced satisfaction with FSM-pricing we ask respondents to answer the question "How satisfied are you with the following?" respondents are asked to rate 14 survey topics. To find out whether respondents feel that FSM-pricing needs to be continued a question is asked to be answered with yes or no: "Should FSM-pricing be continued as an operational practice once the pilot is finalized?" To understand possible reasons behind the closed questions we ask the stakeholders to answer three open questions (max 3 answers are allowed):

1. What is going well during the FSM-pricing pilot that we want to continue?
2. What is not going well during the FSM-pricing pilot that we want to fix?

In order to assess the experienced transparency with regard to project proposals we perform a survey with eight closed questions. The first seven (Q01 to Q07) are intended to find out how respondents experience the quality of artifacts and processes with regard to FSM-pricing. As a response to the question "How would you rate the quality of the following?" respondents are asked to rate these seven survey topics. Next to these questions three additional questions (E01 to E03) are asked: "To what extent did you experience a change on…?" respectively the transparency of proposals during the FSM-pricing pilot, the project cost per FP measured in euros per FP and the project duration per FP measured in days per FP.

RQ3 (To what extent does FSM-pricing help to create cost and time improvements?) is answered by performing quantitative analysis of project data. We analyze the performance of 77 finalized software engineering projects. For our study we use data of three categories of software engineering projects, all performed within COMPANY C:

1. *Repository*: data of historic projects in the period preceding FSM-pricing, not performed by SUPPLIER S (n = 22);
2. *Baseline*: data of finalized projects performed by SUPPLIER S used to prepare the FSM-pricing baseline (n = 16);
3. *Pilot*: data of finalized projects performed during the pilot that are in scope of FSM-pricing (n = 10);
4. *Operational*: data of projects finalized during the six months following the pilot (in scope of FSM-pricing) (n = 29).

In order to benchmark the outcomes of the qualitative analysis with industry peer groups we use a research repository of

331 comparable projects from other companies that we collected in earlier research [31]. All compared peer group projects from this benchmark repository conducted software engineering in business environments. Peer group projects were measured, collected, and recorded in the same way as conducted in this case study.

IV. RESULTS

A. Case and Subject descriptions

In this section we report results based on the three research questions of our study. We sent 41 survey requests by email to 17 employees of COMPANY C and 24 employees of SUPPLIER S. We selected these stakeholders because they are all involved in the FSM-pricing pilot. Twenty seven (27) surveys are returned, of which 2 are assessed to be incomplete (respondents only noted that they knew too little of the subject). 25 surveys are completed (completion rate 61%); the analysis in this study is based on these completed surveys only. TABLE II summarizes the backgrounds of the respondents that completed the survey:

TABLE II BACKGROUNDS FROM SURVEY RESPONDENTS

Respondent background	COMPANY C n=11 (44%)	SUPPLIER S n=14 (56%)
Overall IT-management	28%	29%
FPA Measurement Team	18%	14%
Portfolio Management	27%	0%
Data Warehouse Team	9%	14%
CRM/Billing Team	9%	36%
Other	9%	7%

Besides the results of the survey ratings we collected a large amount of open ended text from our survey. The first open question "What is going well during the FSM-pricing pilot that we want to continue?" resulted in 46 answers. The second open question "What is not going well during the FSM-pricing pilot that we want to fix?" resulted in 47 answers and 44 answers were given to the question "What can we do to improve FSM-pricing?" In total 2,007 words were produced. In this section we label respondents as P1 through P25 and we include results from the open text analysis where applicable. To analyze the free text answers, we adopt the coding technique described by Runeson et al. [10]. We apply high level codes and medium level codes and count the frequency of each code. A summary of the results of this analysis is shown in TABLE I.

B. Results of the Qualitative Analysis

As is common in case studies, answers on surveys contain a substantial element of narrative. As these are representatives of the complexities and contradictions of real life, we include a selection of statements made by the survey respondents in the section on open ended text analysis in our paper. We try to include examples of respondent statements that apply to differences as well as similarities.

TABLE III summarizes the survey results. The two last columns show Effect Size calculated as two measures; 1) for each survey topic the difference between the mean COMPANY C score and the mean SUPPLIER S score, and 2) for each survey topic the difference between the mean Management score (all scores of

TABLE I SUMMARY OF THE OPEN ENDED TEXT ANALYSIS

Category Name / Medium Level Code
Interactions, communications, people
Improved proposal transparency
Improve knowledge of Function Point Analysis and FSM-pricing
Discussion on size when lower price is expected or on waivers
Organization, processes
Uniform, standard and simplified process
Too small projects; no focus on release-based working
Delay due to search for clarity and review
Improve pricing tables (e.g. benchmarking, more realistic figs.)
Promote release-based working based on size
Promote pricing tables based on applications (technology)
Measurements
Perform gap-analysis on FSM-price versus actual effort spent
Requirements
FSM-pricing does not cover non-functional requirements
Low reliability of FSM-pricing when compared to actual effort
Improved Requirement Management
Artifacts
Good quality of Function Point Analysis process and products

respondents with the profile Overall IT-management, FPA Measurement Team, Portfolio Management, and Other) and Development (all scores of respondents with the profile Data Warehouse Team, and CRM/Billing Team). A negative Effect Size indicates COMPANY C / Management respondents are less satisfied with a survey topic than SUPPLIER S / Development respondents. A positive Effect Size indicates COMPANY C / Management respondents are more satisfied with a survey topic than SUPPLIER S / Development respondents.

We found the following with regard to satisfaction with FSM-pricing based on analysis of the survey results:

1) 88% want FSM-pricing as operational practice

On the question "Should FSM-pricing be continued as an operational practice once the pilot is finalized?" 80% answered "Yes"; 8% answered "Ok, but with improvement points (e.g. include effort of non-functional requirements").

2) FPA is appreciated by both parties

Both COMPANY C and SUPPLIER S respondents appreciate the applied FPA method (IFPUG, estimated counts); based upon the highest overall mean score of the survey (3.96). Besides that both parties appreciate the quality of the function point analyses that are performed by SUPPLIER S (3.78), and the reviews done by COMPANY C (3.80).

Qualitative analysis confirmed this finding. Many respondents considered the quality of the FPA high:

Good Function Point review by COMPANY C and SUPPLIER S FPA-teams before proposal submission. (P10)

Appreciate the way Function Point counting is done by SUPPLIER S. (P23)

Many remarks made by respondents were related to requirements; which makes sense since requirements usually are the basis for project proposals. A noteworthy side-effect of FSM-

TABLE III SURVEY RESULTS

Survey Topic (How satisfied are you with the following?)	Nr	Mean Overall	Standard Deviation	Mean Company	Mean Supplier	Effect Size Company/ Supplier	Effect Size Management / Development
Function Point Analysis method (IFPUG, estimated count)	S09	3.96	0.81	4.00	3.92	0.08	0.11
FSM-pricing pilot period itself	S02	3.87	0.55	3.91	3.83	0.08	-0.20
Preparation of the FSM-pricing pilot	S01	3.75	0.90	3.82	3.69	0.13	0.00
Overall FSM-pricing	S15	3.72	0.74	3.64	3.64	0.00	0.08
Advantages of FSM-pricing for COMPANY C	S13	3.68	0.65	3.80	3.58	0.22	-0.30
Pricing table for DWH	S07	3.50	0.73	3.86	3.22	0.63	0.15
Proposal Process (with regard to FSM-pricing)	S12	3.42	0.88	3.70	3.21	0.49	0.06
Management Commitment on FSM-pricing	S04	3.42	0.83	3.64	3.23	0.41	0.25
Advantages of FSM-pricing for SUPPLIER S	S14	3.40	0.68	3.29	3.46	-0.18	0.18
Communication with regard to FSM-pricing	S03	3.39	0.66	3.36	3.42	-0.05	0.22
Setup of the SUPPLIER S Baseline	S06	3.30	0.93	3.55	3.08	0.46	0.13
Pricing table for CRM / Billing	S08	3.28	0.83	3.57	3.09	0.48	0.22
Reliability of the FSM-pricing	S05	3.28	0.94	3.55	3.07	0.47	0.09
Coverage of FSM-pricing	S11	3.26	0.92	2.70	3.69	-0.99	-0.45
Waiver procedure for Function Point Analysis (exclusions)	S10	3.25	1.03	3.00	3.46	-0.46	0.38
Survey Topic (To what extent did you experience change on…?)							
Transparency of Proposals	E01	3.88	0.65	3.82	3.93	-0.11	0.36
Project Cost per FP (Euros per FP)	E02	3.33	0.70	3.40	3.29	0.11	0.17
Project Duration per FP (Days per FP)	E03	3.00	0.76	2.78	3.15	-0.37	0.42
Survey Topic (How would you rate the quality of the following?)							
Function Point Analysis performed by SUPPLIER S	Q02	3.83	0.70	3.70	3.93	-0.23	-0.06
Function Point Analysis Review by COMPANY C	Q03	3.78	0.60	3.73	3.83	-0.11	-0.11
The Overall FSM-pricing method	Q07	3.64	0.57	3.55	3.71	-0.17	-0.22
The SUPPLIER S Proposals based on FSM-pricing	Q06	3.52	0.65	3.55	3.50	0.05	0.12
The CRM / Billing Baseline used for FSM-pricing	Q05	3.47	0.80	3.57	3.40	0.17	-0.05
Requirements delivered by COMPANY C	Q01	3.44	0.65	3.45	3.43	0.03	-0.01
The DWH Baseline used for FSM-pricing	Q04	3.43	0.76	3.71	3.14	0.57	0.55

Sorted by Mean Overall; higher is better.

pricing is that respondents experienced an improvement of the requirement management process during the pilot.

Most of the details are sorted out at the time of proposals. Earlier these details were discussed in design phase. (P17)

The solution is looked into more detail in order to get the right Function Points at the proposal stage itself. This helps in early detection of issues and resolution. (P2)

This positive effect on requirements management might even be one of the main reasons for FSM-pricing success.

3) COMPANY C management: coverage needs improvement

Coverage is about the number of projects in COMPANY C's IT-portfolio that is subject of FSM-pricing. Based on a relatively low mean value for COMPANY C (2.70), combined with an Effect Size of -0.99 between COMPANY C and SUPPLIER S, we conclude that respondents from COMPANY C are more than average dissatisfied about the coverage of FSM-pricing. An Effect Size of -0.45 between Management and Development indicates that coverage is a management rather than a developer concern.

We conjecture a connection with low rating of the waiver procedure by COMPANY C respondents; this procedure allows SUPPLIER S to exclude a project from FSM-pricing. A standard waiver is applied for infrastructure projects, configuration projects, and projects executed by other external suppliers. Also qualitative analysis revealed indications that ongoing discussions tend to be related with waiver requests:

Many ongoing discussions on waiver requests occur. (P20)

4) SUPPLIER S development: reliabiliy needs improvement

In the context of FSM-pricing by reliability we mean whether respondents experience the outcome of FSM-pricing to be in line with their own judgment. SUPPLIER S developers seem dissatisfied with FSM-pricing where it comes to reliability. Proposal process (Effect Size 0.49), both pricing tables (0.48 and 0.63), reliability of FSM-pricing (0.47), and setup of baselines (0.46) are all rated low. We believe these are connected, but we did not find evidence for this in our data.

Looking at this aspect further in the qualitative analysis shows a feeling of disagreement between the outcome of FSM-pricing and effort-based estimates. Many respondents, especially from SUPPLIER S, mention that FSM-pricing does not cover Non-Functional Requirements and complexity (technology).

FPA is not applicable to projects where more testing efforts are required for less development changes. (P5)

All the projects do have different non-functional requirements or technology; due to this the efforts differs. (P2)

The complexity of the changed code does not match with the amount of functionality to be changed, causing a disparity. (P16)

We identified one specific measurement-related issue: the wish to perform a gap-analysis to find any differences between FSM-pricing proposals and actual effort spent in a project:

To keep the counting simple we are considering all the requirements are at average level; we may need to perform gap analysis if the requirements mix is really averaging out on efforts. (P17)

Cross verification with actuals towards the end of project to revalidate the estimates would be an improvement. (P7)

We identified a need for gap-analysis in order to identify differences between (estimated) project cost and actual effort. We consider conducting this gap-analysis as future research.

With regard to the experienced transparency of project proposals we observed one major finding:

5) 84% experienced improved proposal transparency

Many respondents experienced an improvement of the transparency of project proposals during the FSM-pricing pilot (72% said transparency improved; 12% said greatly improved). Qualitative analysis confirmed this finding. Respondents mention improved transparency as a positive outcome of the FSM-pricing pilot:

A good point is that there is less discussion. (P8)

Some respondents see improved transparency as a driver for better requirements or to solve disagreements between customer and supplier:

Instead of plain list of entities that we were maintaining in work-breakdown-structure entities, we now have clarity on what kind of functionality is getting delivered. (P17)

Function points analysis sometimes is a constructive argument in case of disagreement. (P20)

We observed the fact that FSM-pricing is experienced as a uniform, simplified process is on top of respondents' list:

FSM-pricing is a single point for the final estimation, answerable to all stakeholders. The estimation review process becomes very simple. A standardized process, which can be trusted from both vendor and client stakeholders. (P24)

Uniformity in pricing approach as it does not depend on individual components to derive their efforts. (P2)

Avoid delays and budget overruns as estimation can be done at an initial stage against task-based. (P13)

C. Results of the Quantitative Analysis

Data from four categories of 77 software engineering projects are used for quantitative analysis of project data (resp. Repository, Baseline, and Pilot). In TABLE V we summarize the performance indicators for these four project categories. The analysis resulted in the following findings:

1) Project Duration per FP not in sync with peer groups

Analysis of the performance of the software engineering projects of COMPANY C shows that, although the project cost are in line with the prevailing market, the organization suffers from project durations that are substantially longer than those of peer groups in industry. An external benchmark against historic data of 331 finalized software engineering projects [31] from different companies shows that a majority of the finalized projects of COMPANY C are cost effective (average Project Cost per FP is 46% better than the peer groups, see TABLE VI), yet project durations are longer than the average of the total research group (average Project Duration per FP is more than twice that of the peer groups, see TABLE VI). This finding is applicable to all four categories of software projects performed within COMPANY C in our research repository, yet Project Duration per FP is worsening during the pilot.

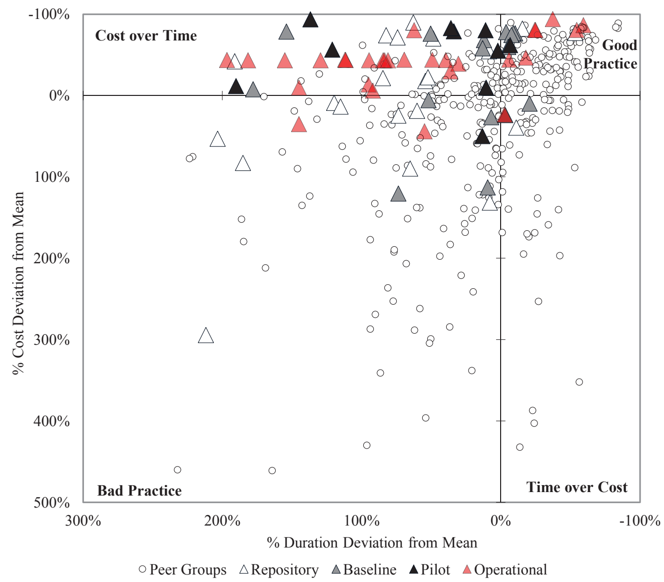

FIGURE 1 COST / DURATION MATRIX

We plot both all COMPANY C and peer group projects in a Cost / Duration Matrix (see FIGURE 1) [31]. This matrix shows for each project the measure of deviation from the average trend line (average of peer group projects plus COMPANY C projects) expressed in a percentage; negative when below the average trend line, positive when above the trend line. The matrix is divided in four quadrants. Each quadrant is characterized by the measure of negative or positive deviation from the average trend. When analyzed it shows that 80% of the projects is assessed to have a longer than average duration. 25% of the projects are in the Bad Practice quadrant; these projects perform in both cost and duration worse than average. 55% ends up in the quadrant Cost over Time; costs are less than average, yet project duration takes longer than average. Due to these deviating percentages we argue that Company A's Project Duration per FP, measured in days per FP, is not in sync with its peer groups; COMPANY C should improve its Project Duration per FP in order to stay competitive in the market.

Our analysis is that the bad Project Duration per FP is caused by two problems. First; the combined release approach of COMPANY C causes waiting time (waste) and unnecessary dependencies between projects. Second; average project duration conform industry, yet combined with small average project size cause a bad Duration per FP as illustrated in the following.

2) Small projects block improvement

A finding with regard to project size is that from 2013-Q3 onwards substantially more very small projects (e.g. projects smaller than 30 FPs) are performed. We did not find any reason that could explain this reduction of project size. Although

TABLE IV PERFORMANCE OVER TIME

Performance Indicator	2012-2013	2014	Delta
Number of projects (n)	38	39	n.a.
Average project Size (FP)	168	68	-59%
Throughput (FP)	6,366	2,660	-29%[1]
Project Cost per FP (EUR/FP)	2,116	1,679	-21%
Project Duration per FP (Days/FP)	2.00	3.52	76%
Average project Duration (Months)	11,69	7,90	-25%

[1]Throughput percentage is calculated based on extrapolation per year.

TABLE V PERFORMANCE OVER FOUR PROJECT CATEGORIES

Performance Indicator	Rp	Bl	Pi	Op
Number of projects (n)	22	16	10	29
Average project Size (FP)	157	183	25	55
Project Cost per FP (EUR/FP)	2,651	1,485	2,560	1,539
Project Duration per FP (Days/FP)	2.35	1.58	7.17	2.95
Average project Duration (Months)	12,11	7,53	7,38	7,67

Rp = Repository, Bl = Baseline, Pi = Pilot, Op = Operational

TABLE VI PERFORMANCE COMPARED TO PEER GROUPS

Performance Indicator	COMPANY C	Peer Gr.	Delta
Number of Projects (n)	26	331	n.a.
Average Project Size (FP)	126	261	-52%
Project Cost per FP (EUR/FP)	1,604	2,983	-46%
Average Project Cost (K Euro)	203K	780K	-74%
Project Duration per FP (Days/FP)	2.20	1.04	112%
Average Project Duration (Months)	9,14	8.92	2%

Performance of Company in comparison with peer group projects from our research repository. Only finalized projects that were performed by SUPPLIER S are incorporated.

smaller projects are from a cost point of view advantageous for SUPPLIER S, portfolio managers of COMPANY C are responsible for the construction of a specific release portfolio (a number of projects combined in one release; to be delivered at one specific moment). The idea that small projects from an economy-of-scale perspective should be combined is mentioned by some respondents in the open ended text as well:

SUPPLIER S divides the offer in small pieces; we must have release based funding to make use of economy-of-scale. (P8)

Too many small projects are negative for COMPANY C due to economy-of-scale effects. (P3)

We observed that in 2014 the throughput (total delivered number FPs) is approximately 29% lower than in the preceding years (see TABLE IV). One can argue that the maybe rather rigid approach of FSM-pricing is not sufficiently encouraging for SUPPLIER S due to a somewhat single-sided focus on cost reduction. However, COMPANY C promotes the idea that delivery of more throughput where applicable is desired. Looked upon from this side FSM-pricing underlines the delivery of more value for less money; and at the same time it rewards throughput enlarging by creating more turnover for the supplier.

3) Cost improves; yet, Duration does not

Looking at cost and duration over time (see TABLE V) we find that Cost per FP (the cost per FP measured over the whole project lifecycle from initiation to technical Go Live) improves by 21% in 2014 onwards compared to the years before. However, Duration per FP is not. Next to our finding that Duration per FP is substantially higher than that of the peer groups, no sustained improvements with regard to project durations are seen when assessed over time. Duration per FP shows a worsening trend. As discussed before the small size of many projects and the amount of waste in projects plays an important role here.

V. DISCUSSION

Analysis with regard to RQ1 (To what extent are both parties involved in the case study satisfied with FSM-pricing?) resulted in four findings. First, 88% of the respondents of our survey want FSM-pricing as an operational practice once the FSM-pilot is finalized. Second, the applied method for FPA, including the counting itself as performed by and SUPPLIER S and the review by COMPANY C, is appreciated highly by both respondents of both parties. Third, coverage of FSM-pricing with regard to COMPANY C's IT-portfolio is experienced as to be improved, mainly by managers from COMPANY C. Additional analysis of the measure of coverage of FSM-pricing with regard to the IT-portfolio shows that at finalization of the FSM-pricing pilot 27% of all IT-portfolio costs were calculated based on FSM-pricing. At the end of the Operational period (end 2014) the coverage was improved to 52%. The remaining 45% is among others related to infrastructure (19%), support (17%), third party projects (5%) and small innovations (3%).

Fourth, developers from SUPPLIER S are dissatisfied with the reliability of FSM-pricing. The major reason for this seems to be that they experience little possibilities to incorporate non-functional requirements and complexity in project proposals. From a statistical point of view all projects are treated as average, where non-functional requirements and related complexity are incorporated in both trend lines. To finalize our discussion on RQ1; an additional positive signal with regard to this is that after evaluation of the FSM-pricing pilot both COMPANY C and SUPPLIER S agreed upon continuation of the approach as an operational practice.

With regard to RQ2 (To what extent does FSM-pricing help to improve transparency of project proposals?) a noteworthy finding was that a large majority (84%) of the respondents of the survey experienced that transparency of project proposals is improved during the FSM-pricing pilot. We observed that the majority of discussions moved from effort (and price) estimate to waiver requests and getting requirements ready for FPA. Noteworthy is that FPA seems to have a positive effect on requirements management.

Looking at RQ3 (To what extent does FSM-pricing help to create cost and time improvements?) quantitative analysis of the performance of the COMPANY C projects taught us that Cost per FP improved during the study, where Duration per FP is not improving over time: this even shows a deterioration. This deterioration however seems to be caused by the fact that average project size gets smaller during the study while average project durations improve notably over time: average project duration in 2014 was even better than that of peer groups in industry.

A. Evaluation of Validity

1) Construct validity

With regard to the degree to which a test measures what it claims to be measuring a remark is in place on FPA. We used functional documentation as a source for FPA; a consequence is that low quality documentation could have led to low quality FPAs, however, we thoroughly reviewed all sets on completeness and correctness. Two (2) out of four (4) FPA specialists were certified; yet, all involved FPA specialists were highly trained and experienced FP-counters. With regard to quality of data we argue that all project data was reviewed by the applicable COMPANY C project manager, all data on project cost was reviewed by the financial controller of COMPANY C, all project data was presented to and discussed with COMPANY C management.

2) Internal validity

We warranted the extent to which a causal conclusion is based on our study, by normalizing all project data with the functional size in FPs. In this way we were able to objectively compare performances of all projects in order to minimize systematic error. Based on the number of software projects, the diversity of projects and business domains within COMPANY C, and the fact that we measured and analyzed software project portfolios as a whole in an empirical way we argue that the effect of outliers is limited and that the risk on bias is mitigated responsibly.

3) External validity

Whether the study results can be generalized to settings outside the study, we argue that due to the limited scope of the performed case study (one sourcing company and one main supplier) it is too early to generalize the above mentioned findings to other companies and suppliers of software projects.

B. Relation to Existing Evidence

From our analysis of related work, it is clear that pricing in itself is a topic that has received little attention from the research community. Yet pricing is a topic of great practical value, which strongly affects the outcome (success or failure) of a software development project. The many budget overruns reported for such projects, may very well be more attributable to inadequate pricing than to poor project execution.

C. Impact/Implications

Our research shows that an evidence-based approach, in which historical data on key performance indicators are used in combination with a simple (power) regression, can lead to prices that are satisfactory to both suppliers and commissioning parties. It emphasizes a holistic approach, in which pricing is considered for the full IT portfolio of an organization, in combination with a supplier in an outsourcing relation. A major prerequisite for this approach is the availability of historical project data. This implies that the approach is only applicable to organizations willing and capable to aim for a long term solution.

The need for historical project data is likely also one of the causes why pricing has received limited attention in the research community; few researchers have access to such data. A way out of this dilemma may be opening up performance data for government-funded projects, making them available for researchers. Besides bringing new research insights, this might also help governments to reach more adequate prices for their IT projects.

D. Limitations

The reader should consider several limitations when interpreting our results. First, the survey has limited generalizability due to the limitation of respondents to 25 stakeholders. Determination of survey topics was done by members of both measurement teams, limited by the length of the survey (10-minutes). Further, the results of the ratings within the survey have to be looked upon with low significance in mind. We did not ask respondents to connect their open ended text data with the answers given in the rating part of the survey.

Second, we conducted the study only within COMPANY C and SUPPLIER S, so the results may not generalize elsewhere. Since we did not find any other study on a comparable single, statistical pricing approach, we cannot predict what the outcome of our method will be in other companies.

Third, our study focused on transparency of proposals and cost and duration improvement. The respondents might have been influenced by this focus and emphasize these aspects in their answers.

VI. CONCLUSIONS AND FUTURE WORK

The key contributions of this paper are:

RQ1: We demonstrate that FSM-pricing is successfully used in practice of COMPANY C and SUPPLIER S, as a statistical, evidence-based pricing approach for software project proposals.

RQ2: We show that using FSM-pricing as a single instrument, without intervention of expert judgment-based opinions, leads in COMPANY C and SUPPLIER S to an improved transparency of project proposals and satisfied stakeholders from both the customer and the supplier.

RQ3: We demonstrate that FSM-pricing does lead to cost improvement within COMPANY C and SUPPLIER S. Cost per FP shows to be in line with external peer groups. Duration per FP on the contrary is too high when benchmarked externally and shows a deteriorating trend, probably caused by the fact that average project size gets smaller over time.

A. Future Work

The research presented opens up a number of avenues for further research. From a benchmarking perspective, our current approach distinguishes between data-warehousing and CRM / Billing projects. Further research is needed to come up with general guidelines on how to group projects into sufficiently cohesive units to permit adequate pricing. Another concern that arose from our case study is dealing with non-functional requirements such as security or infrastructure.

Delivery of smaller software projects in equal project durations seems to result in a lower Duration per FP; however, its needs to be researched whether the amount of value delivered by a project influences such performance perception. With regard to including non-functional requirements it might be interesting to perform future research on possibilities to use IFPUG SNAP (Software Non-functional Assessment Process) besides FPA. Approaches like COCOMO 2 introduce factors to compensate for such project characteristics, but whether this works well in combination with the purely statistical approach investigated in the present paper calls for additional research.

ACKNOWLEDGMENTS

We thank both COMPANY C and SUPPLIER S for their generosity to agree on using project and survey data for our study. Furthermore we thank Philippe Kruchten, Frank Vogelezang, Kim Herzig for their valuable feedback.

References

[1] M. Jørgensen, "A review of studies on expert estimation of software development effort," *The Journal of Systems and Software,* vol. 70, no. IEEE, pp. 37-60, 2004.

[2] B. Boehm, "Software Engineering Economics," *IEEE Transactions on Software Engineering,* vol. 10, no. 1, pp. 7-19, 1984.

[3] K. Moløkken and M. Jørgensen, "A Review of Surveys on Software Effort Estimation," in *IEEE Proceedings of the 2003 International Symposium on Empirical Software Engineering (ISESE'03)*, 2003.

[4] C. Verhoef, "Quantitative IT Portfolio Management," *Elsevier - Science of Computer Programming,* vol. 45, no. 1, pp. 1-96, 2002.

[5] R. Glass, Facts and Fallacies of Software Engineering, Addison Wesley, 2002.

[6] A. Abran, I. Silva and L. Primera, "Field studies using functional size measurement in building estimation models for software maintenance," *Journal of Software Maintenenace and Evolution: Research and Practice,* vol. 14, no. John Wiley & Sons, Ltd., pp. 31-64, 2002.

[7] C. Gencel and O. Demirors, "Functional Size Measurement Revisited," *ACM Transactions on Software Engineering and Methodology,* vol. 17, no. 3, pp. 15:1-15:36, June 2008.

[8] IFPUG, IFPUG FSM Method: ISO/IEC 20926 - Software and systems engineering – Software measurement – IFPUG functional size measurement method, New York: International Function Point User Group (IFPUG), 2009.

[9] F. Heemstra, "Software cost estimation," *Information and Software Technology,* vol. 34, no. 10, pp. 627 - 639, 1992.

[10] P. Runeson, M. Host, A. Rainer and B. Regnell, Case Study Research in Software Engineering; Guidelines and Examples, Hoboken, New Jersey. USA: John Wily & Sons, 2012.

[11] B. Boehm, C. Abts, A. Winsor Brown, S. Chulani, B. K. Clark, E. Horowitz, R. Madachy, D. J. Reifer and B. Steece, Software Cost Estimation with COCOMO II, Englewood Cliffs, NJ: Prentice-Hal, 2000, ISBN 0-13-026692.

[12] L. Putnam and W. Meyers, Five Core Metrics, The Intelligence Behind Succesfull Software Management, New York: Dorset House Publishing, 2003.

[13] L. Fischman, K. McRitchie and D. Galorath, "Inside SEER-SEM," *Crosstalk - The Journal of Defense Software Engineering,* vol. April, pp. 26-28, 2005.

[14] A. L. Lederer and J. Prasad, "Information systems software cost estimating: a current assessment," *Journal of Information Technology,* vol. 8, no. Palgrave Macmillan, pp. 22-33, 1993.

[15] F. Bergeron and J.-Y. St-Arnaud, "Estimation of information systems development efforts: a pilot study," *Information & Management,* vol. 22, no. Elsevier, pp. 239-254, 1992.

[16] C. Dekkers and P. Forselius, "Scope Management: 12 Steps for ICT Program Recovery," *CROSSTALK The Journal of Defense Software Engineering,* vol. January / February, pp. 16-21, 2010.

[17] B. Czarnacka-Chrobot, "Rational pricing of business software systems on the basis of functional size measurement: a case study from Poland," in *Proceedings 7th Software Measurement European Forum (SMEF)*, Rome, Italy, 2010.

[18] T. Menzies and M. Shepperd, "Special issue on repeatable results in software engineering prediction," *Springer Empirical Software Engineering,* vol. 17, no. 1, pp. 1-17, 2012.

[19] M. Jorgensen and T. M. Gruschke, "The impact of lessons-learned sessions on effort estimation and uncertainty assessments," *IEEE Transactions on Software Engineering,* vol. 35, no. 3, pp. 368-383, 2009.

[20] R. Valerdi, "Convergence of expert opinion via the wideband delphi method: An application in cost estimation models," in *Incose International Symposium*, Denver, USA, 2011.

[21] C. Passos, A. Braun, D. Cruzes and M. Mendonca, "Analyzing the impact of beliefs in software project practices," in *IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*, 2011.

[22] E. Kupiainen, M. V. Mäntylä and J. Itkonen, "Using metrics in Agile and Lean Software Development–A systematic literature review of industrial studies," *Information and Software Technology,* vol. 62, pp. 143-163, 2015.

[23] L. Fink and Y. Lichtenstein, "Why project size matters for contract choice in software development outsourcing," *ACM SIGMIS Database,* vol. 45, no. 3, pp. 54-71, 2014.

[24] R. Madachy, W. Rosa, B. Boehm, B. Clark and T. Tan, "Us dod application domain empirical software cost analysis," in *IEEE International Symposium onEmpirical Software Engineering and Measurement (ESEM)*, 2011.

[25] A. Abran, J.-M. Desharnais, M. Zarour and O. Demırörs, "Productivity based software estimation model: an economics perspective and an empirical study," in *9th International Conference on Software Engineering Advances–ICSEA*, 2014.

[26] N. Ramasubbu, M. Cataldo, R. Balan and J. Herbsleb, "Configuring global software teams: a multi-company analysis of project productivity, quality, and profits," in *Proceedings of the 33rd international conference on Software engineering. ACM*, 2011.

[27] N. Ramasubbu and R. K. Balan, "Overcoming the challenges in cost estimation for distributed software projects," in *Proceedings of the 34th International Conference on Software Engineering. IEEE*, 2012.

[28] H. C. Benestad and J. E. Hannay, "A comparison of model-based and judgment-based release planning in incremental software projects," in *Proceedings of the 33rd International Conference on Software Engineering. ACM*, 2011.

[29] A. Albrecht, "Measuring Application Development Productivity," in *Joint Share Guide, and IBM Application Development Symposium 14-17 October 1979*, 1979.

[30] H. Huijgens, G. Gousios and A. v. Deursen, "Pricing via Functional Size: A Case Study of 77 Outsourced Projects - Technical Report TUD-SERG-2014-012," Delft University of Technology, Delft, The Netherlands, 2014.

[31] H. Huijgens, R. v. Solingen and A. v. Deursen, "How To Build a Good Practice Software Project Portfolio?," *ICSE Companion 2014 Companion Proceedings of the 36th International Conference on Software Engineering (SEIP),* vol. 2014, no. IEEE, pp. 64-73, 2014.